

## # Exploratory Data Analysis (EDA) Report on Netflix Dataset

### Introduction

This report presents an Exploratory Data Analysis (EDA) of a Netflix dataset containing **8,790 entries** with details such as show type, title, director, country, release year, rating, duration, and genres.

---

### Dataset Overview

- **Total Entries:** 8,790
- **Columns:** 10
- **Data Types:** Mostly categorical (object) with one numerical (`release_year`).
- **Missing Values:** None (all columns are complete).
- **Duplicates:** None found.

#### ***Columns:***

1. `show_id` – Unique identifier
2. `type` – Movie or TV Show
3. `title` – Show title
4. `director` – Director name
5. `country` – Country of origin
6. `date_added` – Date added to Netflix
7. `release_year` – Release year
8. `rating` – Age/content rating
9. `duration` – Duration of movie (minutes) or TV seasons
10. `listed_in` – Genres/categories

---

### Key Findings

## ***1. Content Type Distribution***

- **Movies:** 6,126
- **TV Shows:** 2,664

➡■ *Movies dominate the Netflix library (~70%).*

---

## ***2. Top Contributing Countries***

Top 10 countries with the most Netflix content:

- United States – 3,240
- India – 1,057
- United Kingdom – 638
- Pakistan – 421
- Canada – 271
- Japan – 259
- South Korea – 214
- France – 213
- Spain – 182
- Turkey – 156

---

## ***3. Release Year Distribution***

- Netflix has content ranging across several decades.
- Peak content production is observed **after 2015**.
- The majority of titles are from **2017–2021**.

---

## ***4. Ratings Distribution***

- Popular ratings: **\*\*TV-MA, TV-14, TV-PG, TV-Y7\*\***
- Indicates Netflix has a wide range of adult and family-friendly content.

---

## **5. Duration Analysis (Movies Only)**

- Converted duration into numeric (minutes).
- Most movies are **\*\*80–120 minutes\*\*** long.
- Few movies exceed 150 minutes.

---

## **6. Genre Distribution**

Top 10 genres (from `listed\_in`):

- Documentaries
- Stand-Up Comedy
- Dramas
- International Movies
- Comedies
- Action & Adventure
- Children & Family Movies
- Romantic TV Shows
- Crime TV Shows
- TV Dramas

---

## **7. Correlation Analysis**

- Label Encoding was applied to categorical features.
- Correlation heatmap shows weak-to-moderate relationships among features.
- Stronger correlation observed between:

- `type` and `rating`
- `release\_year` and `rating`

---

## Visualizations Summary

- **Countplots:** Distribution of Movies vs TV Shows
- **Bar Charts:** Top 10 countries, Top genres
- **Histograms:** Release years, movie durations
- **Boxplots:** Duration vs Ratings
- **Heatmap:** Feature correlations

---

## Conclusion

- Netflix library is **movie-dominated** with strong representation from the **US and India**.
- Content production has grown significantly in recent years.
- The platform offers diverse ratings and genres catering to both **adults and families**.
- Duration patterns suggest Netflix prefers standard-length films (80–120 minutes).